JOURNAL OF MULTIVARIATE ANALYSIS 16, 253–259 (1985)

The Nonnegative MINQUE Estimate

HELENE MASSAM*

York University, Toronto, Canada

AND

JOCHEN MULLER

University of Pittsburgh

Communicated by B. K. Sinha

The purpose of this paper is to give a characterization of the nonnegative MINQUE estimate for variance components. A similar characterization has been given by Pukelsheim but only in some special cases. The proof presented here uses results from convex programming and emphasizes certain geometrical aspects of the nonnegative MINQUE estimate problem.

1. INTRODUCTION

The purpose of this paper is to illustrate the use of some primal convex programming techniques in the field of variance estimation. Let us consider the variance component model

$$\begin{gathered} y = X\beta + e, \qquad \beta \in \mathscr{R}^k, \\ E(e) = 0, \qquad \operatorname{var}(e) = V_\tau = \sum_{i=1}^{m} \tau_i V_i, \qquad \tau \in \theta \subset \mathscr{R}^m. \end{gathered} \tag{1.1}$$

Here X is a known $n \times k$ matrix and V_i are known symmetric positive semi-definite $n \times n$ matrices, while $\beta \in \mathscr{R}^k$ and $\tau = (\tau_1, ..., \tau_m)'$ is a vector of unknown nonnegative parameters, the latter belonging to some subset θ of $\mathscr{R}^m$.

Received January 20, 1983; revised September 24, 1983.
AMS Subject Classifications: Primary 62F10, 62J05.
Key Words and Phrases: variance components, nonnegative MINQUE, convex programming, cones, faces.
* Research partially supported by NSERC Grant A8592.

0047-259X/85 $3.00

Quadratic unbiased estimators of a parametric function $q'\tau$ ($q \in \mathscr{R}^m$) are quadratic forms $y'Ay$, where A is a symmetric matrix satisfying the conditions of unbiasedness:

$$X'AX = 0 \qquad \text{and} \qquad (A, V_\tau) = q'\tau \qquad \text{for} \quad \tau \in \theta. \tag{1.2}$$

Here (A, B) stands for trace(AB). Following Rao [8 or 9], $y'A^*y$ is called a MINQUE (minimum norm quadratic unbiased estimator) if it minimizes some norm $\|A\|$ among all translation invariant unbiased estimators $y'Ay$ of $q'\tau$. However, the MINQUE estimate of a nonnegative $q'\tau$ might be negative. Therefore, special attention has to be directed to nonnegative estimation, for example, if one is interested in estimating a single variance component. For a thorough review on this subject the reader is referred to Pukelsheim (7). In (7), Pukelsheim derived a sufficient—and in special cases also necessary—condition for nonnegative MINQUE estimators. While his approach rests on duality in convex programming, he asked how results from primal theory relate to nonnegative variance estimation.

The present paper answers this question and arrives at a similar condition which is now necessary and sufficient. Moreover, looking at the convex program from a different point of view yields a condition of Lehmann–Scheffé type. To begin with, we outline the necessary tools from convex programming theory.

2. A Characterization of Optimality for a Convex Programming Problem

We shall define a convex programming problem and give necessary and sufficient conditions for a point to be optimal. But, before we do so, we need to define some concepts used in convex analysis. In the following, $\mathscr{X}$ denotes a Hilbert space with inner product (,).

Definition 1. The subset $\mathscr{S}$ of $\mathscr{X}$ is a convex cone if given any x in $\mathscr{S}$, $\lambda x \in \mathscr{S}$ for all $\lambda > 0$ and if given any x_1, x_2 in $\mathscr{S}$, $x_1 + x_2 \in \mathscr{S}$.

Definition 2. Given a nonempty closed convex cone $\mathscr{S}$ in $\mathscr{X}$, a function g from $\mathscr{X}$ to $\mathscr{X}$ is said to be $\mathscr{S}$-convex if for any x_1, x_2 in $\mathscr{X}$, and λ in $(0, 1)$

$$g(\lambda x_1 + (1-\lambda)\, x_2) - \lambda g(x_1) - (1-\lambda)\, g(x_2)\, \varepsilon - \mathscr{S}.$$

It is clear that if a function is linear, then it is convex with respect to any cone $\mathscr{S}$.

DEFINITION 3. The dual of a cone $\mathscr{S}$ is the set

$$\mathscr{S}^+ = \{x' \in \mathscr{X}' : (x', x) \geqslant 0 \text{ for all } x \text{ in } \mathscr{S}\}.$$

The annihilator of a cone $\mathscr{S}$ is the set $\mathscr{S}^\perp = \{x' \in \mathscr{X}' : (x', x) = 0$ for all x in $\mathscr{S}\}$.

DEFINITION 4. A subset $\mathscr{F}$ of a convex cone $\mathscr{S}$ is a face of $\mathscr{S}$ if for any x_1, x_2 in $\mathscr{S}$, $x_1 + x_2 \in \mathscr{F}$ implies $x_1 \in \mathscr{F}$ and $x_2 \in \mathscr{F}$.

Let us now consider the following problem

$$\text{Minimize } p(a) \tag{C}$$

such that $g(a) \in \mathscr{S}$ and $a \in \Omega$, where p is a convex function from $\mathscr{X}$ to $\mathscr{R}$, $\mathscr{S}$ is a closed convex cone in $\mathscr{X}$, g is an $\mathscr{S}$-convex function from $\mathscr{X}$ to $\mathscr{X}$ and Ω is a convex set in $\mathscr{X}$. For this convex programming problem (C), one can define the minimal face and the generalized cone of constancy at a given point a.

DEFINITION 5. The minimal face for a given program (C) is the smallest face of $\mathscr{S}$ containing the image $g(\mathscr{A})$ of the feasible set $\mathscr{A} = g^{-1}(\mathscr{S}) \cap \Omega$ of (C). This minimal face is denoted by $\mathscr{S}^f$.

DEFINITION 6. Let g be the constraint function defined in (C), then the generalized cone of constancy for (C) at a is defined by

$$D_g^=(a) = \{d \in \mathscr{X} : \exists \alpha(d) > 0 \text{ with } g(a + td) \in \mathscr{S}^f - \mathscr{S}^f \text{ for } t \in [0, \alpha(d)]\}.$$

Using these concepts, Borwein and Wolkowicz in [3, Corollary 4.3] (see also Massam [6], in the case where $\Omega = \mathscr{R}^n$) gave the following conditions for a point a^* to be optimal.

Suppose that $\mathscr{S}^+ + (\mathscr{S}^f)^\perp$ is closed, then a^* is optimal for (C) if and only if there exists an b in $\mathscr{S}^+$ such that

$$\nabla p(a^*) \in b \nabla g(a^*) + (D_g^=(a^*) \cap \operatorname{cone}(\Omega - a^*))^+$$

and

$$(b, g(a^*)) = 0,$$

where cone $(\Omega - a^*)$ denotes the convex cone generated by $\Omega - a^*$.

In the special case where g is a linear function, $b\nabla g(a^*)$ becomes $g(b)$ and $D_g^=(a^*) = \{d \in \mathscr{X} : g(d) \in \mathscr{S}^f - \mathscr{S}^f\} = g^{-1}(\mathscr{S}^f - \mathscr{S}^f)$. Then $(D_g^=(a^*))^+ = (D_g^=(a^*))^\perp = (g^{-1}(\mathscr{S}^f - \mathscr{S}^f))^\perp = g(\mathscr{S}^{f\perp})$ (for a proof of this last result, see Drygas [4, p. 33], and so $(D_g^=(a^*) \cap \operatorname{cone}(\Omega - a^*))^+ =$

$g(\mathscr{S}^{f\perp}) + \text{cone}(\Omega - a^*)^+$. It then follows from Corollary 4.3 in (3) that $a^* \in A$ is optimal for (C) if and only if

$$\nabla p(a^*) = g(s) + g(c) + u \tag{2.1}$$

for some $u \in (\text{cone}(\Omega - a^*))^+$, $c \in (\mathscr{S}^f)^\perp$, and $s \in \mathscr{S}^+$ with $(s, a^*) = 0$.

3. The Nonnegative MINQUE

The space $\mathscr{X} = \text{sym}\,\mathscr{R}^n$ of all real symmetric $n \times n$ matrices forms a Hilbert space with the inner product $(A, B) = \text{trace}(AB)$. We write $A \geqslant 0$ to indicate that A belongs to the cone of nonnegative semi-definite matrices. As follows from (1.2) a nonnegative unbiased estimator $y'Ay$ of some parametric function $q'\tau$ is characterized by

$$A \geqslant 0, \qquad AX = 0, \qquad (A, V_\tau) = q'\tau \qquad \text{for} \quad \tau \in \theta. \tag{3.1}$$

Such an estimator is always translation invariant and the condition $AX = 0$ can be written $MAM = A$, where $M = I - XX^+$ is the projection onto the null space of X'. Thus (3.1) is equivalent to

$$A \geqslant 0, \qquad MAM = A, \qquad (A, MV_\tau M) = q'\tau \qquad (\tau \in \theta). \tag{3.2}$$

The quadratic form $y'A^*y$ is called nonnegative MINQUE if it minimizes the norm $\|A\| = (A, A)^{1/2}$ among all A satisfying (3.2). Let $\mathscr{S}$ denote the pointed closed convex cone of all matrices $A \geqslant 0$ and let Ω be the linear manifold of all matrices A meeting $(A, MV_\tau M) = q'\tau$ $(\tau \in \theta)$. We then have the convex program

$$\min \tfrac{1}{2} \|A\|^2 \qquad \text{such that } A = MAM \in \mathscr{S}, \qquad A \in \Omega. \tag{3.3}$$

Instead of this we consider

$$\min \tfrac{1}{2} \|A\|^2 \qquad \text{such that } MAM \in \mathscr{S}, \qquad A \in \Omega. \tag{3.4}$$

Denoting the set of all feasible A of this program by $\mathscr{A}$ the feasible set of (3.3) reads $M\mathscr{A}M$. The following lemma ensures that the solutions of both problems coincide.

Lemma. *Let $\mathscr{A} \subset \mathscr{X}$ be such that $M\mathscr{A}M \subset \mathscr{A}$. A^* minimizes $\|A\|$ such that $A \in \mathscr{A}$ if and only if A^* equals MA^*M and minimizes $\|A\|$ such that $A \in M\mathscr{A}M$.*

The proof of the lemma follows immediately from the inequality $\|MAM\| \leqslant \|A\|$ which in turn is valid since the mapping $A \to MAM$ is an orthogonal procjection in sym $\mathscr{R}^n$.

In order to utilize the result of the previous section for our problem (3.4) we have to determine $\mathscr{S}^+$, $\mathscr{S}^f$, and $(\text{cone}(\Omega - A^*))^+$. Obviously $\mathscr{S}^+$ equals $\mathscr{S}$. Further, writing $\mathscr{R}(A)$ for the image of a linear mapping A, any face $\mathscr{S}_0$ of $\mathscr{S}$ must be of the form

$$\mathscr{S}_0 = N\mathscr{S}N = \{A \in \mathscr{S} : \mathscr{R}(A) \subset \mathscr{R}(N)\}$$

for some projection N in $\mathscr{S}$ (cf. Bellisard, Iochum, and Lima (1)). Now $\mathscr{S}^f$ is the smallest face containing $M\mathscr{A}M$, and $M\mathscr{A}M$ represents the set of all nonnegative unbiased estimators of $q'\tau$. Thus

$$\mathscr{S}^f = \{A \in \mathscr{S} : \mathscr{R}(A) \subset \mathscr{R}(\bar{A})\},$$

where $\bar{A} = M\bar{A}M$ is a nonnegative unbiased estimator of $q'\tau$ with maximum image. It alsways exists, for if there is some $A \in M\mathscr{A}M$ whose image is not contained in that of $\bar{A}$ replace $\bar{A}$ by $\frac{1}{2}(A + \bar{A})$. This procedure stops not later than the nth step. Denoting the projection onto $\mathscr{R}(\bar{A})$ by N we obtain

$$\mathscr{S}^f = N\mathscr{S}N, \qquad \text{where} \quad \mathscr{R}(N) \subset \mathscr{R}(M),$$

since $\mathscr{S}^f$ is included in $M\mathscr{S}M$.

The cone $\mathscr{S}^f$ generates the subspace $\mathscr{S}^f - \mathscr{S}^f = N(\mathscr{S} - \mathscr{S})\,N = N(\text{sym}\,\mathscr{R}^n)\,N$. Its orthogonal complement is

$$(\mathscr{S}^f)^\perp = \{C : NCN = 0\}.$$

As is readily verified $\mathscr{S} + (\mathscr{S}^f)^\perp$ is the subset of all A such that NAN is nonnegative. So $\mathscr{S} + (\mathscr{S}^f)^\perp$ must be closed for it is the inverse image of the closed set $\mathscr{S}$ under the continuous mapping $A \to NAN$.

Finally we need $(\text{cone}(\Omega - A^*))^+$. Because A^* belongs to Ω we have

$$\begin{aligned}
\Omega - A^* &= \{A - A^* : (A, MV_\tau M) = (A^*, MV_\tau M),\ \tau \in \theta\} \\
&= \{D : (D, MV_\tau M) = 0 \text{ for } \tau \in \theta\} \\
&= (M\mathscr{V}M)^\perp = \text{cone}(\Omega - A^*),
\end{aligned}$$

where $\mathscr{V} = \{V_\tau : \tau \in \theta\}$ is the set of all covariance matrices of the model (1.1). Therefore $\Omega - A^*$ is a subspace and its dual equals its orthogonal complement.

$$(\Omega - A^*)^+ = (\Omega - A^*)^\perp = (M\mathscr{V}M)^{\perp\perp} = \text{span}\, M\mathscr{V}M.$$

We now can apply the result of Section 2. Since $\nabla \frac{1}{2}\, \|A\|^2 = A$ we obtain from (2.1), $A^* \in \mathscr{A}$ is a solution of the convex program in (3.4) if and only if

$$A^* = MSM + MCM + MV_\lambda M$$

for some $\lambda \in \operatorname{span} \theta$, C with $NCN = 0$, and $S \in \mathscr{S}$ with $(S, A^*) = 0$. As A^* is an element of $\mathscr{S}^f = N\mathscr{S}N$ we have $NA^*N = A^*$. Moreover, $\mathscr{R}(N)$ is contained in $\mathscr{R}(M)$ and so $NM = N = MN$. Thus

$$A^* = NA^*N = NSN + NV_\lambda N. \tag{3.5}$$

Here $NSN \geqslant 0$ and $(NSN, A^*) = (S, NA^*N) = (S, A^*) = 0$.

Let us look now at the minimization problem from a different point of view. Instead of considering the program (3.4) we can as well deal with

$$\min \tfrac{1}{2} \|A\|^2, \qquad A \in \mathscr{S}, \quad A \in \Omega. \tag{3.6}$$

Again the above lemma ensures that the solution A^* of this problem coincides with that of the original program. Let $\tilde{g} = 0$ and apply the results of Section 2 to

$$\min \tfrac{1}{2} \|A\|^2, \qquad A \in \tilde{\Omega} = \mathscr{S} \cap \Omega.$$

It then yields the condition

$$A^* \in (\tilde{\Omega} - A^*)^+$$

which is equivalent to

$$(A^*, D) \geqslant 0 \qquad \text{for all } D \in (M\mathscr{V}M)^\perp \tag{3.7}$$

such that $A^* + D \geqslant 0$.

We summarize our results (3.5) and (3.7) in a proposition.

PROPOSITION. *A nonnegative unbiased estimator $y'A^*y$ of $q'\tau$ in the model* (1.1) *is a nonnegative MINQUE if and only if one of the two equivalent conditions holds true*:

(i) $(A^*, D) \geqslant 0$ *for all* $D \in (M\mathscr{V}M)^\perp$ *such that* $A^* + D \geqslant 0$.

(ii) $A^* = S + NV_\lambda N$ *for some* $\lambda \in \operatorname{span} \theta$ *and* $S \geqslant 0$ *such that* $(S, A^*) = 0$.

Here $\mathscr{V} = \{V_\tau : \tau \in \theta\}$, $M = I - XX^+$, *and* N *is the projection onto* $\mathscr{R}(\bar{A})$, *where* $y'\bar{A}y$ *is a nonnegative unbiased estimator of* $q'\tau$ *with maximum image.*

Condition (i) above was already derived in Pukelsheim's (7) by the aid of another theorem of convex analysis. It can be easily seen. One just has to look at $A = A^* + \lambda D$ and

$$\|A\|^2 = \|A^*\|^2 + \lambda^2 \|D\|^2 + 2\lambda(A^*, D) \geqslant \|A^*\|^2$$

for $0 \leqslant \lambda \leqslant 1$. Because this is analogous to the method Lehmann and Scheffé [5] used for unrestricted estimation one might call (i) the Lehmann–Scheffé condition for nonnegative MINQUE. Moreover, (i) is easily deduced from (ii) observing that $D = NDN = MDM$ when $D \in (M\mathscr{V}M)^{\perp}$ and $A^* + D \geqslant 0$. Then

$$(A^*, D) = (A^* - NV_{\lambda}N, D) = (S, D) = (S, A^* + D) \geqslant 0.$$

In this way the sufficiency of condition (ii) is a simple consequence of (i). Conversely, in order to prove the necessity of (ii) some convex analysis technique seems to be inevitable.

References

[1] Bellisard, J. Iochum, B., and Lima, R. (1976). Homogeneous and facially homogeneous self-dual cones. *Linear Algebra Appl.* **19** 1–16.

[2] Borwein, J., and Wolkowicz, H. (1981). Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A* **30** 369–381.

[3] Borwein, J., and Wolkowicz, H. (1981). Characterization of optimality for the abstract convex program with finite dimensional range. *J. Austral. Math. Soc. Ser. A* **30** 390–412.

[4] Drygas, H. (1970). "The Coordinate Free Approach to Gauss–Markov Estimation," Lecture Notes in Operations Research and Mathematical Systems, Springer-Verlag, Berlin.

[5] Lehmann, E. L., and Scheffé, H. (1950). Completeness, similar regions and unbiased estimation, Part I. *Sankhya,* **10** 305–340.

[6] Massam, H. (1979). Optimality conditions for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A* **27** 141–162.

[7] Pukelsheim, F. (1981). Linear models and convex geometry aspects of nonnegative variance estimation. *Mathematische Operationsforch. Statist. Ser. Statist.* **12** 271–286.

[8] Rao, C. Radhakrishna (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **65** 161–172.

[9] Rao, C. Radhakrishna (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115.